

# **NIH PROGRAM EVALUATION GUIDE:**

HOW TO DEVELOP A PROPOSAL FOR  
ONE PERCENT EVALUATION SET-ASIDE FUNDING

**JANUARY 2, 2001**

**OFFICE OF EVALUATION, OFFICE OF SCIENCE POLICY  
OFFICE OF THE DIRECTOR, NATIONAL INSTITUTES OF HEALTH**

This page was left blank intentionally.

## ACKNOWLEDGMENTS

---

We greatly appreciate the time and effort of program administrators, contracting and evaluation staff, and other NIH personnel who contributed to this document. Specifically, we would like to thank the members of the Technical Merit Review Committee (Lura Abbott, Jim Corrigan, Jared Jobe, Cherie Nichols, Jim Onken, Barbara Perrone, Carl Roth, and Chuck Sherman) who provided guidance and clarity regarding their expectations for evaluation proposals; Miriam Keltz, Mona Rowe, Elliot Siegel, Judith Whalen, and Fred Wood, who reviewed this document from a proposer's perspective; Jonathan Ferguson, who provided useful input regarding developing an evaluation budget; and Joan Bailey, Mark Boroush, Robin Butcher, and Janet Healy, who provided on-going conceptual and technical support.

# TABLE OF CONTENTS

---

<b>Preface.....</b>	<b>vi</b>
<b>Introduction.....</b>	<b>vii</b>
<b>Section 1: Program to be Evaluated.....</b>	<b>1</b>
1.1 Cover Page .....	1
1.2 Program to be Evaluated.....	1
1.3 Program Goals .....	1
<b>Section 2: Need for an Evaluation.....</b>	<b>3</b>
2.1 Type of Evaluation.....	3
2.2 Purpose of the Evaluation .....	3
2.3 Use of Results .....	3
2.4 Review of the Literature .....	4
2.5 Timeliness of the Evaluation .....	4
<b>Section 3: Evaluation Design .....</b>	<b>5</b>
3.1 Study Questions .....	5
3.2 Target Population.....	5
3.3 Key Variables.....	5
3.4 Conceptual Framework.....	6
<b>Section 4: Data Collection and Analysis .....</b>	<b>7</b>
4.1 Data Sources .....	7
4.2 Data Collection Strategies.....	7
4.3 New Data Collection Instruments.....	8
4.4 Clearance Requirements .....	8
4.5 Data Integrity .....	9
4.6 Ethical Considerations .....	9
4.7 Data Preparation.....	9
4.8 Data Analysis .....	9
<b>Section 5: Evaluation Results .....</b>	<b>10</b>
5.1 Products of the Evaluation .....	10
5.2 Dissemination of Results .....	10
<b>Section 6: Project Management.....</b>	<b>11</b>
6.1 Project Implementation.....	11

## TABLE OF CONTENTS, Continued

---

6.2 Advisory Committee .....	11
6.3 Estimated Timeline for the Evaluation .....	11
<b>Section 7: Budget Estimate .....</b>	<b>12</b>
7.1 Estimated Cost .....	12
7.2 Anticipated Funding Sources .....	12
 <b>Appendix A: Types of Program Evaluations .....</b>	 <b>13</b>
<b>Appendix B: Program Goals, Performance Measures, and Comparison Measures .....</b>	<b>15</b>
<b>Appendix C: Examples of Study Questions .....</b>	<b>18</b>
<b>Appendix D: Examples of Conceptual Frameworks .....</b>	<b>21</b>
<b>Appendix E: Developing an Evaluation Budget Estimate .....</b>	<b>25</b>
<b>Bibliography .....</b>	<b>28</b>
<b>Glossary .....</b>	<b>32</b>

## PREFACE

---

Program evaluation is a critical tool for identifying the processes that contribute to a program's success or failure and for measuring the extent to which NIH programs are accomplishing their goals. The purpose of this guide is to improve the planning and implementation of NIH-supported program evaluations. It was developed principally for program staff who are seeking NIH One Percent Evaluation Set-Aside funds to evaluate eligible NIH programs, but it should prove to be useful to others as well.

This guide is intended to provide 1) a roadmap for NIH staff responsible for preparing a proposal for NIH One Percent Evaluation Set-Aside funds, and 2) key evaluation concepts and definitions useful in reviewing and overseeing evaluation projects. It was designed to be useful to a broad audience, ranging from staff with little or no experience in program evaluation to staff having extensive evaluation expertise. Those with little experience in planning or conducting a program evaluation should read this guide carefully, being sure to follow the proposal preparation instructions systematically. Those with more experience, particularly those who have successfully applied for NIH One Percent Evaluation Set-Aside funding, should also benefit from the straightforward approach used and may note that proposal requirements have been clarified.

The Office of Evaluation in NIH's Office of Science Policy, Office of the Director, developed this program evaluation guide. Diane Buckley, an Operations Research Analyst in the Office of Evaluation, provided principal direction, and Dr. Marcia Carlyn of Carlyn Consulting contributed significantly to all aspects of the guide.

As program staff in the NIH Institutes, Centers and Offices gain more experience in developing program evaluation proposals, additional issues and key questions will likely emerge. Accordingly, this guide is expected to be a working document, subject to future refinement to meet the needs of its users. An electronic version is available at [http://www3.od.nih.gov/ospde/ep\\_review.htm](http://www3.od.nih.gov/ospde/ep_review.htm). If you have questions or comments, please contact Diane Buckley or John Uzzell at (301) 496-9285.

/s/

John K. Uzzell  
Director, Office of Evaluation, Office of Science Policy,  
Office of the Director, NIH

# INTRODUCTION

---

NIH seeks to uncover new knowledge about the prevention, detection, diagnosis, and treatment of disease and disability. To that end, NIH supports a wide range of programs related to the conduct of medical research, the training and professional development of scientists, and the construction and maintenance of the laboratory facilities needed to support cutting-edge research. One important way that NIH administrators determine the extent to which these varied programs are operating efficiently and achieving their intended effects is by conducting program evaluations.

Program evaluations are systematic investigations or studies that involve assessing the worth and/or performance of particular programs. In most cases, the underlying purpose of a program evaluation is to help NIH administrators improve a program or make other programmatic decisions (e.g., how to allocate resources). For the purposes of evaluation, a “program” is broadly defined as any set of activities funded by NIH to achieve one or more predefined goals (also referred to as “program goals”).

NIH’s Institutes and Centers (ICs) frequently use program funds to support evaluation activities to improve decision-making and, ultimately, enhance program performance. However, many of NIH’s activities are crosscutting in nature and require trans-NIH program evaluations (i.e., program evaluations that involve more than one IC) to be examined effectively. A key source of funding for trans-NIH program evaluations is the One Percent Evaluation Set-Aside.<sup>1</sup>

## *One Percent Evaluation Set-Aside*

Public Law 91-296, passed in 1970, authorizes the Secretary of the Department of Health and Human Services (DHHS) to use up to one percent of its program appropriations authorized under the Public Health Service (PHS) Act for program evaluations and evaluation-related activities (see 42 USC Sec. 238(j)). This “One Percent Evaluation Set-Aside” fund is available to NIH and other PHS agencies within DHHS to assess the effectiveness of federal health programs and to identify ways to improve program implementation.

---

<sup>1</sup> If you are seeking support for an IC-specific program evaluation or another evaluation-related activity, see the Office of Evaluation website ([http://www3.od.nih.gov/ospde/ep\\_review.htm](http://www3.od.nih.gov/ospde/ep_review.htm)) or contact the Office of Evaluation for guidance.

Administration of the NIH portion of the One Percent Evaluation Set-Aside is the responsibility of the Office of Evaluation, Office of Science Policy (OSP), within the Office of the Director (OD). This responsibility is facilitated by two trans-NIH committees:

The *Evaluation Policy and Oversight Committee* (EPOC) determines the size of the NIH One Percent Evaluation Set-Aside budget, identifies and directs improvements in the evaluation set-aside program, reviews funding requests for conceptual merit and policy relevance, and makes funding decisions. The EPOC consists of IC representatives at the level of Director and Deputy Director and OD representatives at the level of Associate Director. The Associate Director for Science Policy, OD serves as the EPOC Chairperson.

The *Technical Merit Review Committee* (TMRC) reviews all funding requests for eligibility and technical merit and makes recommendations to the EPOC regarding the funding of requests. The TMRC consists of IC and OD representatives who have evaluation expertise and are knowledgeable about NIH programs. The Director of the Office of Evaluation serves as the TMRC Chairperson.

### ***Eligibility Guidelines***

Four types of program evaluations are eligible for NIH One Percent Evaluation Set-Aside funding (see Appendix A for additional information):

*Needs Assessment* – A program evaluation aimed at systematically determining the nature and extent of the problems that a proposed or existing program should address.

*Feasibility Study* – A systematic assessment of the optimal approach for evaluating a program, including which evaluation designs and data collection strategies can and should be used. This type of study is sometimes called an evaluability assessment.

*Process Evaluation* – A systematic assessment of program operations to determine whether a program is being conducted as planned, whether expected output is being produced, and/or how program-critical processes can be improved.

*Outcome Evaluation* – A systematic assessment of program accomplishments and effects to determine the extent to which a program's intermediate and/or long-term goals have been achieved.



### ***Preparing an Evaluation Proposal***

Sections 1-7 of this guide provide specific instructions for preparing an evaluation proposal for NIH One Percent Evaluation Set-Aside funding. The following topics are covered and *must* be addressed in your proposal:

- Program to be Evaluated
- Need for an Evaluation
- Evaluation Design
- Data Collection and Analysis
- Evaluation Results
- Project Management
- Budget Estimate

In addition, Appendices A-E provide supporting materials, including typical evaluation strategies used for each type of program evaluation; examples of program goals, performance measures, comparison measures, study questions and conceptual frameworks; and tips on how to develop an evaluation budget estimate. Finally, a detailed glossary presents definitions and examples of key terms used throughout the guide.

### ***Review of Proposals***

The review procedures vary according to the amount of One Percent Evaluation Set-Aside funds requested and the complexity of the proposed evaluation. For information on the length of the review process and for the most current review schedule, contact the Office of Evaluation or your IC Evaluation Officer or visit the Office of Evaluation website at <http://www1.od.nih.gov/osp/de/>.

### ***Application Procedure***

To request One Percent Evaluation Set-Aside funds to support an eligible program evaluation, send an electronic version (either a Microsoft Word or WordPerfect document) of the complete evaluation proposal to the Office of Evaluation via [evaluate@od.nih.gov](mailto:evaluate@od.nih.gov). Proposals should not exceed 15 pages in length and must have at least 12 point type with one inch margins.

# SECTION 1

## PROGRAM TO BE EVALUATED

---

### QUESTIONS TO BE ADDRESSED IN SECTION 1

- [1.1](#) Is the cover page of the proposal complete?
- [1.2](#) Which NIH program is the focus of the evaluation?
- [1.3](#) What are the goals of the program to be evaluated?

#### 1.1 Cover Page

Is the cover page of the proposal complete? It should include the:

- Title of the evaluation
- NIH Institute or Center (IC) that will assume primary responsibility for the evaluation
- Contact information for the person submitting the proposal (i.e., name, title, telephone number, fax number, mailing address, and email address)
- ICs or government agencies that are co-sponsoring the evaluation

#### 1.2 Program to be Evaluated

Which NIH program is the focus of the proposed evaluation? For the purposes of NIH One Percent Evaluation Set-Aside Funds, a program is broadly defined as a set of activities funded by NIH to achieve one or more predefined goals (referred to as “program goals”).

Briefly describe the program to be evaluated, including its organizational location within NIH and the fiscal year it was established (or is likely to be established). Indicate the approximate size of the program in terms of dollars, full-time-equivalent personnel (FTEs), and/or facilities.

#### 1.3 Program Goals

What are the documented goals of the program to be evaluated (or the proposed goals if the program is not yet established)? Indicate which of these goals are relevant to this evaluation. If program goals will be developed as part of a needs assessment, identify any program goals that have been proposed; otherwise indicate why no program goals have been specified in the proposal.

Program goals are the intended effects of a program, as noted in authorizing legislation or other documents written when the program was established. In some cases, additional program goals that are not listed in official documents may be included in the evaluation. For a program that is not yet established, the program goals should summarize the anticipated effects of the new program. There are three types of program goals:

- *Process goals* – Goals that describe how the program should operate and what levels of output should be expected.
- *Intermediate goals* – Goals that describe specific outcomes the program should achieve in the near term.
- *Long-term goals* – Goals that describe the ultimate outcomes the program is designed to achieve.

The evaluation does not need to include all three types of program goals. See the Glossary and Appendix B for more information on program goals, including examples.

## SECTION 2

# NEED FOR AN EVALUATION

---

### QUESTIONS TO BE ADDRESSED IN SECTION 2

- [2.1](#) What type of evaluation is planned?
- [2.2](#) What is the primary purpose of the evaluation?
- [2.3](#) How will the evaluation results be used?
- [2.4](#) Was a formal literature review or an informal review of related studies conducted?
- [2.5](#) What is the rationale for conducting the evaluation at this time?

#### 2.1 Type of Evaluation

What type of evaluation is planned? Specify which of the following best describes the proposed evaluation:

- Needs assessment
- Feasibility study
- Process evaluation
- Outcome evaluation

The four major types of program evaluations are described in Appendix A.

#### 2.2 Purpose of the Evaluation

What is the primary purpose of the proposed evaluation? Briefly describe the main objective(s) in conducting the evaluation. If the proposed study is part of a multi-phase evaluation project, identify the phase of the proposed study. Also state the primary purpose of each of the other phases, including any anticipated future phases.

#### 2.3 Use of Results

How will the evaluation results be used? Briefly describe the different types of stakeholders and explain why they are likely to be interested in, be affected by, or use the findings of the study. Also describe how the sponsoring IC or OD office, other ICs or OD offices, and/or other government agencies are likely to use the results of the evaluation, and identify any factors that could have an impact on the usefulness of the results.

**2.4 Review of the Literature (if applicable)**

Was a formal literature review or an informal review of related studies conducted? If so, briefly describe the review conducted and any findings relevant to the present evaluation. If not, explain why a literature review was not conducted.

See [http://www1.od.nih.gov/osp/de/ep\\_es.htm](http://www1.od.nih.gov/osp/de/ep_es.htm) for a listing of ongoing and recently completed evaluations that have been supported with NIH One Percent Evaluation Set-Aside Funding.

**2.5 Timeliness of the Evaluation**

What is the rationale for conducting the evaluation at this time? Explain any compelling reason for prompt action (e.g., a Congressional mandate, Executive Order, national commission report, or comparable policy document).

## SECTION 3

# EVALUATION DESIGN

---

The evaluation design along with the data collection and analysis plan (Section 4) describe the overall “blueprint” for evaluating the program.

### QUESTIONS TO BE ADDRESSED IN SECTION 3

- [3.1](#) What are the key questions that the evaluation must answer?
- [3.2](#) What is the primary group about which information is needed?
- [3.3](#) What specific information is needed to answer the study questions?
- [3.4](#) Has a conceptual framework (or logic model) been developed?

### 3.1 Study Questions

What are the key questions that the evaluation must answer? List the most important study questions (e.g., 2 to 5 key questions) as well as any hypotheses that will be tested to answer the study questions. Examples of study questions are presented in Appendix C.

### 3.2 Target Population

What is the primary group or groups about which information is needed to answer the study questions? A target population may consist of one or more groups of individuals or of objects (e.g., grant awards, academic institutions) having certain characteristics. Describe the target population, including its approximate size, general characteristics, and any subgroups within the population that will be studied. If applicable, identify the individual unit within the target population for which data will be collected and analyzed. Additional information on a target population and unit of analysis may be found in the Glossary.

### 3.3 Key Variables

What specific information is needed to answer the study questions? List the *most important variables* for which data will be collected, categorizing them as follows:

- Program resources – variables that describe the amount of funding, human capital (e.g., FTEs), infrastructure, and/or other resources allocated to the program or specific program components during a given time period.

- Population characteristics – variables (such as demographic characteristics or types of grants) that describe differences among the members of the target population that may be related to program success.
- Program activities – variables that describe the operations, processes, or other activities that are essential to the program.
- Program goals, performance measures, and comparison measures – interrelated variables that focus on the program’s output and/or outcome. For each program goal used in the evaluation (previously described in Section 1.3), at least one performance measure and corresponding comparison measure should be defined. Detailed definitions of these terms, including examples, are presented in Appendix B.
- External factors – variables that describe conditions or circumstances beyond the control of the program that may influence program success.
- Other variables of interest – any other variables for which data will be collected.

The evaluation does not need to include all of the types of variables listed above and the variable list may be revised during the course of the evaluation.

### **3.4 Conceptual Framework (if applicable)**

Has a conceptual framework (or logic model) been developed to illustrate how the program is intended to work to achieve its goals? If so, include a diagram showing the conceptual framework. Examples of conceptual frameworks are presented in Appendix D.

## SECTION 4

# DATA COLLECTION AND ANALYSIS

---

A variety of strategies may be used in program evaluations to systematically collect and analyze the data needed to answer the study questions (see Appendix A for an overview of typical strategies used for different types of evaluations).

### QUESTIONS TO BE ADDRESSED IN SECTION 4

- [4.1](#) What data sources will be used to obtain information on the key variables?
- [4.2](#) What major data collection strategies will be used to answer the study questions?
- [4.3](#) Will any new data collection instruments be developed?
- [4.4](#) Is any special permission needed before collecting certain data?
- [4.5](#) What strategies will be used to ensure that accurate and complete data are collected?
- [4.6](#) How will the needs/sensitivities of the respondents/program personnel be addressed?
- [4.7](#) How will the data be prepared for analysis?
- [4.8](#) What type of analysis will be used to answer each study question?

#### 4.1 Data Sources

What data sources will be used to provide information on the key variables in the evaluation design? Briefly describe each data source and categorize as follows:

- *Archival data* – Information previously collected for another purpose (also referred to as secondary data).
- *New data* – Information that will be collected specifically for the evaluation (also referred to as primary data).

#### 4.2 Data Collection Strategies

What are the major data collection strategies that will be used to answer the study questions? For each study question, briefly describe how the data that are needed to answer the question will be collected. Include a description of each data collection instrument that will be used, including any questionnaires, interview guides, focus group discussion guides, coding sheets, or other forms for collecting data.



Also explain the processes that will be used to collect the data, including how the particular respondents, documents, records, observations, or other data elements will be selected. If sampling procedures will be employed, briefly describe each type of sampling strategy to be used, including estimated sample sizes and response rates. If any comparison (or control) groups will be used, also describe how the respondents or other data elements within these groups will be selected.

For *archival data*, commonly used data collection strategies include:

- Document reviews
- Database extractions
- Web site reviews
- Literature reviews

If *new data* are needed to answer one or more study questions, commonly used data collection strategies include:

- Personal interviews
- Focus groups
- Expert panels
- Questionnaires or other data collection forms to be completed
- Adding evaluation questions to general-purpose surveys
- Structured observations of program processes

#### **4.3 New Data Collection Instruments (if applicable)**

Will any *new* data collection instruments be developed to answer one or more study questions? If yes, for each new data collection instrument to be developed, briefly describe its key features:

- Primary purpose of the instrument
- The process that will be used to design and pretest the instrument
- How the instrument is expected to be administered
- Approximate number of questions or items
- Any other materials that may accompany the instrument (e.g., instructions, cover letter, postage-paid return envelope)
- Study question(s) to be addressed using the instrument

#### **4.4 Clearance Requirements (if applicable)**

Is special permission needed before collecting certain data? For example, before collecting the same information from ten or more nonfederal employees, Paperwork Reduction Act requirements must be met and clearance must be obtained from the U.S. Office of Management and Budget (OMB). Other data collection efforts require compliance with Privacy Act requirements and/or Institutional Review Board (IRB) approval. If special permission is needed, briefly describe how the clearance requirements will be met. See the

Glossary for more information on Paperwork Reduction Act requirements, Privacy Act requirements, and IRB approval.

#### **4.5 Data Integrity**

What strategies will be used to ensure that the data collected are as accurate and complete as possible? Briefly describe any steps that will be taken to enhance the reliability and validity of the data collected (e.g., pilot tests of instruments and procedures, inter-rater reliability checks). If data collectors will be used, describe how they will be trained and monitored.

#### **4.6 Ethical Considerations**

How will the needs and sensitivities of the respondents and/or program personnel be addressed? Briefly describe any steps that will be taken to provide assurances of confidentiality, to safeguard the security of responses and computerized files, and to minimize the burden placed on respondents and program personnel.

#### **4.7 Data Preparation**

How will the data be prepared for analysis? Briefly describe any verification, quality control, coding procedures, or other steps that will be taken to prepare the data for analysis.

#### **4.8 Data Analysis**

What type of analysis will be used to answer each study question? Briefly describe each planned analysis and categorize it as follows:

- *Descriptive statistics* – used to tabulate, depict, and describe collections of data.
- *Inferential statistics* – used to make inferences about a population by examining a sample from that population.
- *Qualitative analysis* – used to examine data that are in the form of words rather than numbers (e.g., data from focus groups, personal interviews, observations).

See the Glossary for additional information and examples of these types of analyses.

## SECTION 5

# EVALUATION RESULTS

---

### QUESTIONS TO BE ADDRESSED IN SECTION 5

- [5.1](#) What reports and/or other products are planned?
- [5.2](#) How will the findings and any other products of the evaluation be disseminated?

#### 5.1 Products of the Evaluation

What reports and/or other products are planned? Briefly describe the primary purpose of each report/product.

#### 5.2 Dissemination of Results

How will the findings and any other products of the evaluation be disseminated? Briefly describe the intended audience and planned procedures for distributing the information produced, including any web sites that will be used to disseminate the results of the evaluation.

## SECTION 6

# PROJECT MANAGEMENT

---

### QUESTIONS TO BE ADDRESSED IN SECTION 6

- [6.1](#) How will the evaluation be conducted?
- [6.2](#) Will an advisory committee be used?
- [6.3](#) What is the proposed timeline for conducting the evaluation?

#### 6.1 **Project Implementation**

How will the evaluation be conducted (e.g., by NIH staff or by an independent contractor or consultant)? If an independent contractor or consultant will conduct the study, describe the process that will be used to select the contractor/consultant. In all cases, describe the type of expertise needed to conduct the evaluation effectively.

#### 6.2 **Advisory Committee (if applicable)**

Will an advisory committee be used? If yes, briefly describe the committee's responsibilities, its approximate size, the number of meetings that are planned, and the expertise of the individuals who will serve on the committee.

#### 6.3 **Estimated Timeline for the Evaluation**

What is the proposed timeline for conducting the evaluation? Describe or illustrate when each major task is expected to be performed, including (if needed) the time required to select a contractor/consultant and obtain special permission to collect certain data (e.g., OMB clearance).

## SECTION 7

# BUDGET ESTIMATE

---

### QUESTIONS TO BE ADDRESSED IN SECTION 7

- [7.1](#) What is the estimated cost of the evaluation?
- [7.2](#) What funding sources are expected to be used?

#### 7.1 Estimated Cost

What is the estimated cost of the evaluation? A detailed budget estimate should be included in the evaluation proposal when more than \$100,000 in One Percent Evaluation Set-Aside funds are requested. Guidance on how to develop an evaluation budget estimate and an example of an evaluation budget are presented in Appendix E.

#### 7.2 Anticipated Funding Sources

What funding sources are expected to be used? If the evaluation will span multiple fiscal years, indicate the estimated costs and funding sources for each fiscal year as well as for the entire project period. The following format should be used:

Fiscal Year	Estimated Cost	Estimated Amount from Each Funding Source		
		One Percent Evaluation Set-Aside Funds	IC Funds	Other Funds
<b>PROJECT TOTAL</b>				

## APPENDIX A

# TYPES OF PROGRAM EVALUATIONS

Type of Evaluation	Purposes	Typical Evaluation Strategies
<b>Needs Assessment</b>	<p><i>Determine the nature and extent of the problems that a proposed or existing program should address.</i></p> <p>Assess the needs of different stakeholders (e.g., program participants, NIH administrators). For proposed programs, develop appropriate program goals. Determine how a program should be designed or modified to achieve those goals.</p>	<p>Collect and analyze data from document reviews (including program records and literature reviews), databases, in-person and telephone interviews, focus groups, expert panels, structured observations, and/or questionnaires.</p> <p>Use results of data analyses to develop, revise, and/or prioritize program specifications and program goals.</p>
<b>Feasibility Study</b>  <i>Sometimes called an evaluability assessment</i>	<p><i>Determine whether conducting an evaluation is appropriate, design a process evaluation or outcome evaluation for a proposed or existing program, and/or determine whether the evaluation can be conducted at a reasonable cost.</i></p> <p>Determine the optimal approach for evaluating a program. Assess which evaluation designs and data collection strategies can or should be used.</p> <p>Often serves as a Phase 1 evaluation, designed to prepare for a full-scale (Phase 2) outcome evaluation. May include the development of computerized information systems (e.g., databases) as one component of the evaluation design.</p>	<p>Develop data collection instruments (e.g., interview guides, questionnaires, usability tests, coding systems, computerized information systems).</p> <p>Design and conduct pilot tests of data collection instruments and procedures. Design and conduct a pilot study of a small sample of units within the target population.</p> <p>Collect and analyze data from literature and document reviews, in-person and telephone interviews, focus groups, expert panels, structured observations, and/or questionnaires.</p> <p>Develop the final evaluation design (including the type of evaluation, program goals, performance measures, and study questions to be answered) based on the results of the data analyses.</p> <p>Determine if the proposed evaluation can be conducted at a reasonable cost.</p>

Type of Evaluation	Purposes	Typical Evaluation Strategies
<b>Process Evaluation</b>	<p><i>Assess program operations.</i></p> <p>Determine whether a program is being conducted as planned, whether expected output is being produced, and/or how program-critical processes can be improved. Assess the extent to which process goals have been achieved.</p> <p>Examples of <i>process goals</i> include adherence to a pre-established timeline and budget, an increased level of program activities, and a reduction in unit costs.</p>	<p>Collect and analyze data from computerized information systems (particularly internal databases), literature and document reviews, in-person and telephone interviews, structured observations, and/or questionnaires.</p> <p>Answer specific study questions using the data collected to determine if the program is functioning as intended and whether the program operates:</p> <ol style="list-style-type: none"> <li>1) Significantly better (or worse) than in the past,</li> <li>2) Significantly better (or worse) than a comparable program, and/or</li> <li>3) In accordance with recognized standards of performance.</li> </ol>
<b>Outcome Evaluation</b>	<p><i>Assess program effects.</i></p> <p>Determine program accomplishments and effects (specifically whether a program is making progress and/or has fulfilled its goals). Examine the relationship between program activities and their effects, both intended and unintended, to identify why some program variations or strategies worked better than others.</p> <p>Assess the extent to which the program's intermediate and/or long-term goals have been achieved. Examples of <i>intermediate goals</i> include increased publications in peer-reviewed journals and more individuals obtaining doctoral degrees in health-related sciences. An example of a <i>long-term goal</i> is the discovery of a new treatment for a specific disease.</p>	<p>Collect and analyze data from computerized information systems, literature and document reviews, in-person and telephone interviews, structured observations, and/or questionnaires.</p> <p>Answer specific study questions using the data collected, conducting statistical analyses whenever possible, to determine if intermediate and/or long-term program performance is:</p> <ol style="list-style-type: none"> <li>1) Significantly better (or worse) than in the past,</li> <li>2) Significantly better (or worse) than that of a comparable program or control group, and/or</li> <li>3) In accordance with recognized standards of performance.</li> </ol>

## APPENDIX B

# PROGRAM GOALS, PERFORMANCE MEASURES, AND COMPARISON MEASURES

---

Program goals, performance measures, and comparison measures are key variables in program evaluations. Detailed definitions for these terms, including definitions for the different types of program goals, performance measures, and comparison measures are included in this appendix. Several examples are also provided to illustrate how these variables are interrelated.

**Program goals** are the intended effects of a program, as noted in authorizing legislation or other documents written when the program was established. For a program that is not yet established, the proposed program goals should summarize the anticipated effects of the new program (also see Section 1.3). There are three types of program goals:

- Process goals, which describe how the program should operate and what levels of output should be expected.
- Intermediate goals, which describe specific outcomes the program should achieve in the near term.
- Long-term goals, which describe the ultimate outcomes the program is designed to achieve.

**Performance measures** are measurements of program performance that are typically expressed for a given time period. There are three types of performance measures:

- Output measures, which focus on the number of program activities conducted or products produced.
- Outcome measures, which focus on the intermediate and/or long-term accomplishments and effects of the program.
- Efficiency measures, which focus on the cost (in terms of dollars, FTEs, employee-hours, facilities, or other resources) per unit of output or outcome. Efficiency measures are sometimes called cost-effectiveness measures.



**Comparison measures** are measurements against which the performance of the program will be compared. There are three types of comparison measures:

- Measures of the program's prior performance (e.g., baseline performance).
- Measures of a comparable program's performance or a control group's performance.
- Recognized standards of performance.

The extent to which a particular program goal has been achieved should be assessed using one or more performance measures, each with a corresponding comparison measure. For the sake of simplicity, the following examples include only one performance measure and comparison measure for each program goal:

**Example 1**

- Program goal:** Conduct productive grant reviews.
- Performance measure:** Number of grants reviewed by NIH scientific review groups (output measure).
- Comparison measure:** Number of grants reviewed by NSF scientific review groups (comparable program's performance).

**Example 2**

- Program goal:** Provide training opportunities for program participants.
- Performance measure:** Minimum number of workshops held per year (output measure).
- Comparison measure:** At least four workshops should be held each year (recognized standard of performance).

**Example 3**

- Program goal:** Increase citations by NIH-supported investigators.
- Performance measure:** Average number of citations per investigator (outcome measure).
- Comparison measure:** Average number of citations per investigator before the NIH program was implemented (prior performance).

**Example 4**

- Program goal:** Be responsive to a particular need of program participants.
- Performance measure:** Percent of program participants reporting that the need has been adequately addressed (outcome measure).
- Comparison measure:** At least 95% of program participants should report that the need has been adequately addressed (recognized standard of performance).

**Example 5**

- Program goal:** Decrease smoking among youth aged 12-15.
- Performance measure:** Percent of youth aged 12-15 who reported smoking at least 30 cigarettes in July 2000 (outcome measure).
- Comparison measure:** Percent of youth aged 12-15 who reported smoking at least 30 cigarettes in July 1997 (prior performance).

**Example 6**

- Program goal:** Improve the timeliness of website development.
- Performance measure:** Average number of weeks required to develop a new website (efficiency measure).
- Comparison measure:** Average number of weeks required in 1999 to develop a new website (prior performance).

**Example 7**

- Program goal:** Improve the efficiency of program operations.
- Performance measure:** Average cost per hotline call (efficiency measure).
- Comparison measure:** Average cost per hotline call three years ago (prior performance).

## APPENDIX C

# EXAMPLES OF STUDY QUESTIONS

---

Study questions are the key questions that the evaluation is designed to answer. By definition, each type of evaluation typically seeks to answer certain kinds of study questions (see Appendix A for more information on types of evaluations).

***Needs Assessment***      *Study questions typically focus on the nature and extent of the problems that a proposed or existing program should address, as shown by the following examples:*

- What problem or need is the program attempting to address?
- Whom does this program serve? To what extent are their needs being addressed?
- If defined needs are not being addressed, how could the program be revised to address these needs? Would it be more effective to create a new program or revise the existing program?
- What should be the documented goals of a new or revised program?
- For each program goal, what would be a reasonable standard of performance to achieve by a certain year (e.g., 2004)?

***Feasibility Study***      *Study questions typically focus on determining the optimal approach for evaluating a program, as shown by the following examples:*

- Is there adequate justification to conduct a large-scale outcome evaluation at this time? If so, what is the most appropriate approach to use in evaluating the effects of the program (or for regularly monitoring the program's progress)?
- Has the program been operating long enough to have a measurable effect?
- Is the estimated cost of a proposed evaluation reasonable given the cost of the program?
- What performance measures will reveal whether or not the program goals are being achieved?

- Is information available on the program's prior performance that could be used to see if performance has improved?
- Are there recognized standards of performance that could be used to assess success?
- Are there comparable programs for which comparison measures could be obtained?
- What existing data sources (i.e., archival data) could be used to evaluate the program? What new data need to be collected?
- What is the best way to collect evaluation data without imposing an excessive burden on program staff or the public?

***Process Evaluation***

*Study questions typically focus on the extent to which process goals have been achieved, as shown by the following examples:*

- Is the program being implemented as planned (e.g., in terms of the number of different types of activities conducted per month or the type of participants receiving services)? If not, how and why is it being implemented differently?
- During the most recent fiscal year, were all major program activities conducted on time and within budget? If not, how could program processes be improved?
- Has the program succeeded in achieving recognized standards of performance?
- Is the average cost per unit (e.g., per hotline call) significantly higher than it was in the past (e.g., two years ago)? If so, why?

***Outcome Evaluation***

*Study questions typically focus on the extent to which a program's intermediate and/or long-term goals have been achieved, as shown by the following examples:*

- After a certain period of time (e.g., five years), to what extent did the participants in the NIH program achieve the program's intermediate goals? To what extent did they achieve the program's long-term goals?

- Is the program's current performance different from what it was at a certain time in the past (e.g., last year)?
- Have the NIH program participants been more successful than the participants in a comparable non-NIH program? In what ways?
- Which participant characteristics are most related to success?
- Which program activities are most related to success?
- What are the effects (intended and unintended) of the program? Are they positive or negative?

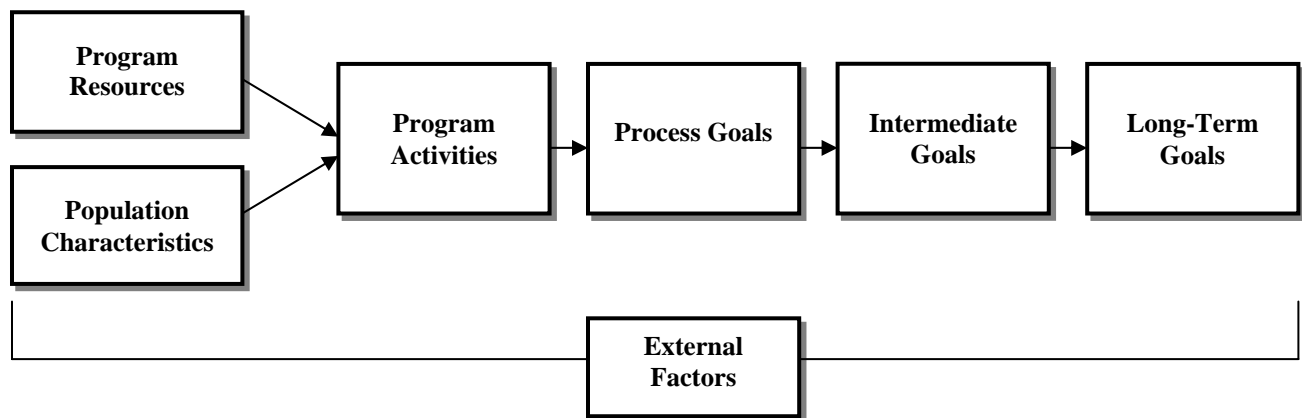
## APPENDIX D

# EXAMPLES OF CONCEPTUAL FRAMEWORKS

---

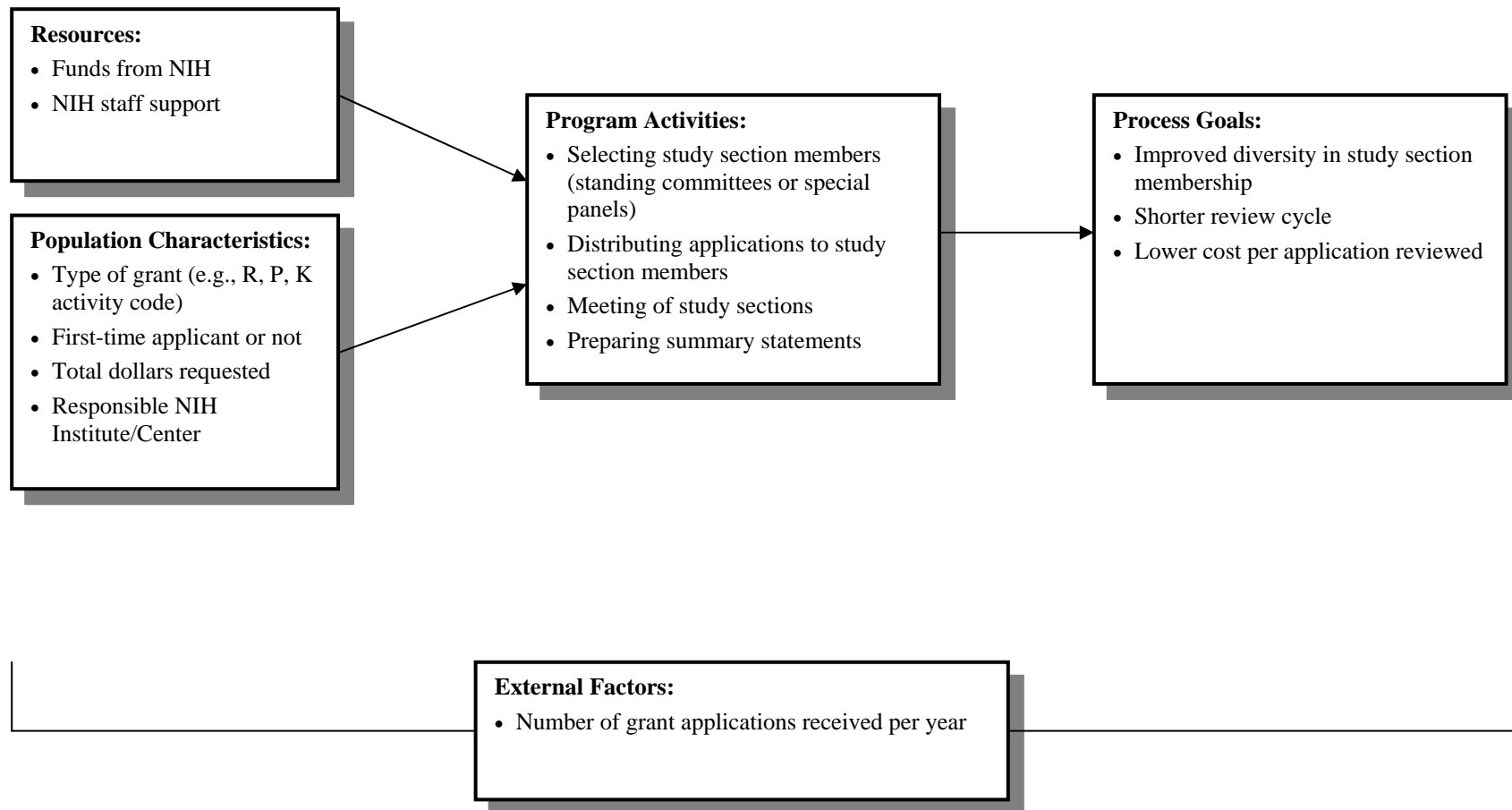
A conceptual framework describes, usually in the form of a diagram, how a particular program is intended to work. Specifically, it demonstrates how program resources, population characteristics, program activities, and external factors (if any) are expected to influence the achievement of a program's specific process, intermediate, and/or long-term goals. A conceptual framework (also called a logic model) may be simple or elaborate, and it can be developed using either common sense or a specific theory as a foundation.

The following is a generic conceptual framework for an outcome evaluation:

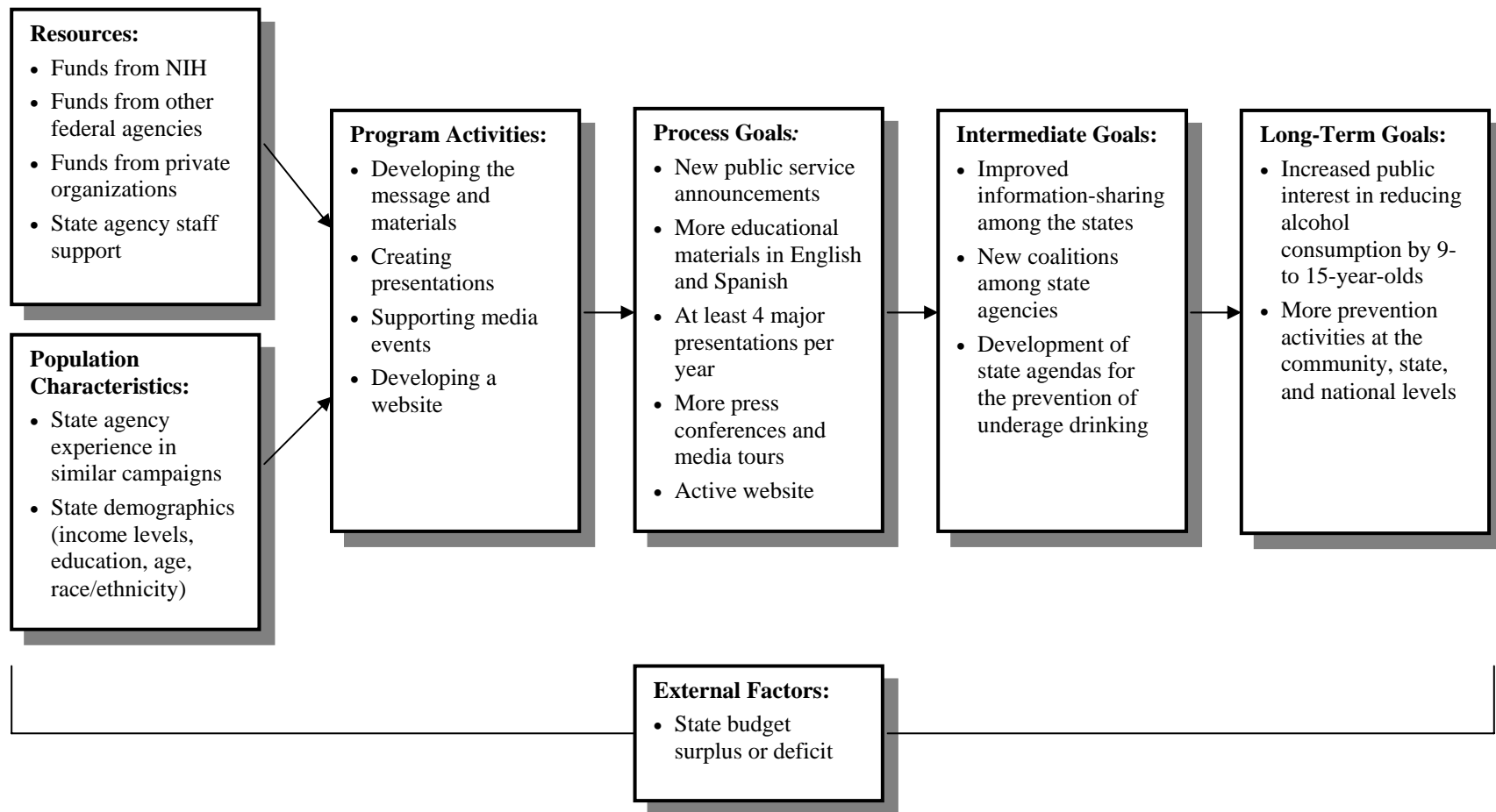


Three examples are shown on the following pages.

A conceptual framework is presented below for an initiative designed to improve the NIH peer-review process. The target population for this process evaluation consists of competitive grant applications reviewed by NIH study sections:

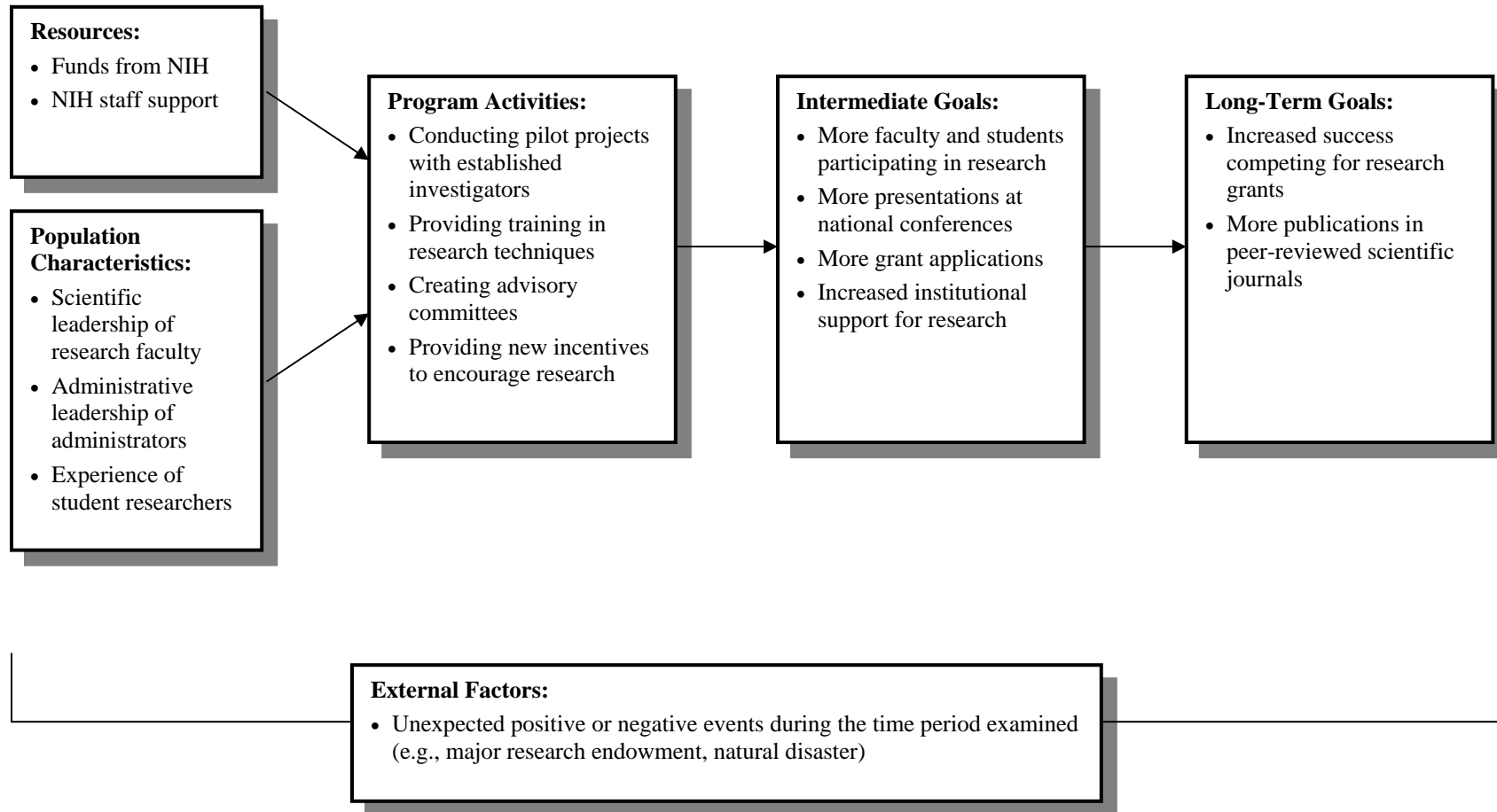


A conceptual framework is presented below for a national public awareness campaign to reduce alcohol consumption by children aged 9-15. The target population for this outcome evaluation consists of state agencies involved in addressing alcohol-related problems:





A conceptual framework is presented below for a cooperative agreement program to enhance the research capacity of minority institutions. The target population for this outcome evaluation consists of research faculty, administrators, and students involved in neuroscience research:



## APPENDIX E

# DEVELOPING AN EVALUATION BUDGET ESTIMATE

---

The budget estimate<sup>2</sup> should provide a detailed picture of the resources needed to conduct the evaluation and the anticipated costs. Specifically, the budget estimate should include the total costs for direct labor, other direct costs, indirect costs, and fee, as applicable to the proposed evaluation. In many cases, the budget will reflect the costs that an independent contractor is expected to incur when conducting the evaluation for the sponsoring IC or OD office.

Below are general guidelines to help in the preparation of an evaluation budget estimate. An example of an evaluation budget is presented at the end of this appendix. However, the figures used in the example are for illustrative purposes only. Contact your IC or OD contracting officer for specific cost estimates for developing your evaluation budget.

### *Estimating Direct Labor Costs*

First, determine the various skills (or labor mix) required to conduct the evaluation. For example, how many senior researchers are needed? How many junior analysts? What about statisticians, information specialists, writers, etc.? Then, determine the hourly rate and total hours of effort required for each labor category. A common approach is to divide the estimated annual salary for each labor category by 2,080 hours to determine each category's hourly rate.

### *Estimating Other Direct Costs*

Consultant costs – If external consultants are needed, the total cost of their services should be estimated using the approach that was recommended for direct labor costs. Consultant labor rates are often higher than direct labor rates because of the special expertise of consultants.

Subcontract costs – If significant costs will be incurred by using subcontractors to perform specific tasks, a cost estimate for each subcontract should be developed. Subcontracts may be necessary if the evaluation involves specialized skills and capabilities that are not available from the organization conducting the study. For example, a subcontractor might be used for analysis of patient data. In most cases, the estimate for subcontract costs is listed in the budget as a lump-sum amount. If the estimate can be broken down into its cost elements (e.g., direct labor, meetings/travel), use that approach.

---

<sup>2</sup> Proposals for over \$100,000 in One Percent Evaluation Set-Aside funds must include an evaluation budget estimate. Proposals valued equal to or less than \$100,000 are not required to include a budget estimate; however, reviewers may request specific budgetary information during the review process.

Meetings/travel costs – If the evaluation design includes site visits, meetings of an expert panel or advisory committee, and/or other events involving local or long-distance travel, determine the total number of events and the duration, location, and expected number of participants for each event. Next, develop an estimate of the average cost of each event, based on the anticipated costs for travel, per diem, and honoraria. Then, determine the total cost per site visit and/or meeting, and multiply this sum by the number of events planned. See the General Services Administration's Travel Management Policy website <http://policyworks.gov/org/main/mt/homepage/mtt/mtthp.htm> for information about federal travel rates.

Miscellaneous supplies/services costs – Additional minor expenses may be incurred during the course of the evaluation for office materials and supplies (e.g., paper, pens, postage) and routine services (e.g., copy and delivery services). For most evaluation projects, the estimated cost of miscellaneous supplies and services is minimal, rarely exceeding 5 percent of direct labor costs.

### ***Estimating Indirect Costs***

Indirect costs are expenses that are difficult to assign to specific project functions. They typically include fringe benefits for the individuals performing direct labor, overhead costs, and general and administrative (G&A) expenses (see the Glossary for definitions of these terms). Indirect costs should be stated as a fixed percentage of total direct costs.

### ***Estimating the Fee***

The fee (or profit) is the dollar amount over and above allowable costs that is to be paid to the organization responsible for conducting the evaluation. The complexity of the task, the level of risk to the organization performing the work, and other factors determine the fee, which is usually presented as a percentage of total estimated costs.

## EXAMPLE OF AN EVALUATION BUDGET<sup>3</sup>

### DIRECT LABOR COSTS

Labor Category	Hourly Rate	Hours	Amount	Total
Project Director	\$ 45	110	\$ 4,950	
Senior Research Associate	\$ 32	400	\$ 12,800	
Information Specialist	\$ 23	300	\$ 6,900	
Writer-Editor	\$ 21	100	\$ 2,100	
TOTAL DIRECT LABOR COSTS		910	\$ 26,750	\$ 26,750

### OTHER DIRECT COSTS

#### Consultants

One consultant with evaluation expertise	\$ 100	60	\$ 6,000
--	--------	----	----------

#### Subcontracts

One subcontract for publication of final report			\$ 5,000
---	--	--	----------

#### Meetings/Travel

Travel costs per meeting (3 people x \$500 fare)			\$ 1,500
Per diem costs per meeting (3 people x \$165 per diem x 2 days)			\$ 990
Honoraria per meeting (3 people x \$200 x 2 days)			\$ 1,200
Cost per meeting (sum of travel, per diem, and fees)			\$ 3,690
Total meetings/travel (\$3,366 per meeting x 2 meetings)			\$ 7,380

#### Miscellaneous Supplies/Services

			\$ 1,000
TOTAL OTHER DIRECT COSTS		\$ 19,380	\$ 19,380

TOTAL DIRECT COSTS	\$ 46,130
--------------------	-----------

### INDIRECT COSTS

Fringe Benefits/Overhead/G&A (90% of Total Direct Costs)	\$ 41,517
--	-----------

TOTAL ESTIMATED COST	\$ 87,647
----------------------	-----------

FEE (8.5% of Total Estimated Cost)	7,450
------------------------------------	-------

TOTAL ESTIMATED COST PLUS FEE	\$ 95,097
-------------------------------	-----------

<sup>3</sup> The figures provided in this example are for illustrative purposes only. Labor categories, hourly rates, and other cost estimates for the evaluation depend on many factors such as the expertise needed to conduct the evaluation and the location of meetings or site visits. Contact your IC or OD contracting officer for specific cost estimates for your planned evaluation.

## BIBLIOGRAPHY

---

- Administration on Children, Youth, and Families. *The Program Manager's Guide to Evaluation*. <http://www2.acf.dhhs.gov/programs/CORE/dox/progman.html>. Last accessed December 2000.
- Blackerby, P. (1996). *Strategic Planning*. Austin, TX: Performance Strategies, Inc.
- Bond, S. L., Boyd, S. E., & Rapp, K. A. (1997). *Taking Stock: A Practical Guide to Evaluating Your Own Programs*. Chapel Hill, NC: Horizon Research, Inc.
- Bureau of Justice Assistance Evaluation Web Site. <http://www.bja.evaluationwebsite.org/index.html>. Last accessed December 2000.
- Burroughs, C. (2000). *Measuring the Difference: Guide to Planning and Evaluating Health Information Outreach*. National Network of Libraries of Medicine, Pacific Northwest Region, Bethesda, MD: National Library of Medicine.
- Burt, M. R., Harrell, A. V., Newmark, L. C., Aron, L.Y., & Jacobs, L.K. (1997). *Evaluation Guidebook: Projects Funded by S.T.O.P. Formula Grants under the Violence Against Women Act*. Washington, DC: The Urban Institute.
- Center for Disease Control and Prevention. (1999). *Framework for Program Evaluation in Public Health*. Morbidity and Mortality Weekly Report, Vol. 48 (No. RR-11). Atlanta, GA. <http://www.cdc.gov/eval/framework.htm>. Last accessed December 2000.
- Center for Disease Control and Prevention. (2000). *FY 2000 One Percent Evaluation Program Proposal Submission Guidelines, Proposal Format, and Scoring Criteria for Proposals*. Office of Program Planning and Evaluation. Atlanta, GA.
- Churchill, G. A. (1987). *Marketing Research: Methodological Foundations* (4<sup>th</sup> ed.). Chicago, IL: The Dryden Press.
- Committee on Science, Engineering, and Public Policy. (1999). *Evaluating Federal Research Programs: Research and the Government Performance and Results Act*. Washington, DC: National Academy Press.
- Electronic Resources for Evaluators. <http://www.nnfr.org/parented/links.html>. Last accessed December 2000.
- Fowler, F. J. (1993). *Survey Research Methods* (2<sup>nd</sup> ed.). Newbury Park, CA: Sage Publications.
- Frechting, J., & Sharp, L. (eds). (1997). *User-Friendly Handbook for Mixed Method Evaluations*. Arlington, VA: National Science Foundation.

- Harrell, A., Burt, M., Hatry, H.P., Rossman, S., Roth, J., & Sabol, W. (2000). *Evaluation Strategies for Human Services Programs: A Guide for Policymakers and Providers*. Washington, DC: The Urban Institute.  
[http://www.bja.evaluationwebsite.org/html/documents/evaluation\\_strategies.html](http://www.bja.evaluationwebsite.org/html/documents/evaluation_strategies.html). Last accessed December 2000.
- Hatry, H. P. (1999). *Performance Measurement: Getting Results*. Washington, D.C.: The Urban Institute Press.
- Hatry, H. P., Blair, L., Fisk, D., & Kimmel, W. (1987). *Program Analysis for State and Local Governments* (2<sup>nd</sup> ed.). Washington, DC: The Urban Institute.
- Hatry, H. P., Fountain, J. R., Jr., Sullivan, J. M. & Kremer, L. (eds.). (1990). *Service Efforts and Accomplishments Reporting: Its Time Has Come*. Norwalk, CT: Governmental Accounting Standards Board.
- Health Communication Unit. (1998). *Evaluating Health Promotion Programs*. Toronto, Ontario: Centre for Health Promotion, University of Toronto.
- Institute of Medicine. (1998). *Scientific Opportunities and Public Needs: Improving Priority Setting and Public Input at the National Institutes of Health*. Washington, DC: National Academy Press.
- Isaac, S., & Michael, W. B. (1997). *Handbook in Research and Evaluation: A Collection of Principles, Methods, and Strategies Useful in the Planning, Design, and Evaluation of Studies in Education and Behavioral Sciences* (3<sup>rd</sup> ed.). San Diego, CA: EdITS.
- Joint Committee on Standards for Educational Evaluation. (1994). *The Program Evaluation Standards* (2nd ed.). Thousand Oaks, CA: Sage Publications.
- King, J.A., Morris, L. L., & Fitz-Gibbon, C. T. (1987). *How to Assess Program Implementation*. Newbury Park, CA: Sage Publications.
- Kosecoff, J. & Fink, A. (1982). *Evaluation Basics: A Practitioner's Manual*. Beverly Hills, CA: Sage Publications.
- McKillip, J. (1987). *Need Analysis: Tools for the Human Services and Education*. Newbury Park, CA: Sage Publications.
- McNamara, C. (1998). *Basic Guide to Program Evaluation*.  
[http://www.mapnp.org/library/evaluatn/fnl\\_eval.htm](http://www.mapnp.org/library/evaluatn/fnl_eval.htm). Last accessed December 2000.
- Moore, D. S. (1991). *Statistics: Concepts and Controversies* (3rd ed.). New York, NY: W.H. Freeman and Company.
- Morris, L. L., Fitz-Gibbon, C. T., & Freeman, M. E. (1987). *How to Communicate Evaluation Findings*. Newbury Park, CA: Sage Publications.
- Morris, L. L., Fitz-Gibbon, C. T., & Lindheim, E. (1987). *How to Measure Performance and Use Tests*. Newbury Park, CA: Sage Publications.
- Muraskin, L.D. (1993). *Understanding Evaluation: The Way to Better Prevention Programs*. Rockville, MD: Westat. <http://www.ed.gov/PDFDocs/handbook.pdf> Last accessed December 2000.

- National Institutes of Health. (February, 2000). *Government Performance and Results Act: FY 2001 Performance Plan and FY 1999 Performance Report*. Bethesda, MD.
- Office of Management and Budget. (1998). *Overview of Strategic Plans, Annual Performance Plans, and Annual Program Performance Reports*. OMB Circular No. A-11. Washington, DC.
- Office of Management and Budget. (February, 1995). *Primer on Performance Measurement*. Washington, DC. <http://www.npr.gov/library/resource/gpraprmr.html>. Last accessed December 2000.
- Patton, M.Q. (1987). *How to Use Qualitative Methods in Evaluation*. Newbury Park, CA: Sage Publications.
- Patton, M. Q. (1990). *Qualitative Evaluation and Research Methods* (2<sup>nd</sup> ed.). Newbury Park, CA: Sage Publications.
- Porteous, N. L., Sheldrick, B. J., & Stewart, P. J. *A Program Evaluation Toolkit*. Ottawa, Ontario: Ottawa-Carleton Health Department. <http://www.uottawa.ca/academic/med/epid/intro.htm>. Last accessed December 2000.
- Rossi, P. H., & Freeman, H. E. (1993). *Evaluation: A Systematic Approach* (5th ed.). Newbury Park, CA: Sage Publications.
- Rouda, R. H. & Kusy, M. E. (1995). *Needs Assessment: The First Step*. [http://www.alumni.caltech.edu/~rouda/T2\\_NA.html](http://www.alumni.caltech.edu/~rouda/T2_NA.html). Last accessed December 2000.
- Rowe, Mona. (April, 1990). *National Institute of Allergy and Infectious Diseases (NIAID) Evaluation Guide*. Bethesda, MD: National Institutes of Health.
- Stecher, B. M., & Davis, W. A. (1987). *How to Focus an Evaluation*. Newbury Park, CA: Sage Publications.
- Stevens, F., Lawrenz, F. & Sharp, L. (1993). *User-Friendly Handbook for Project Evaluation: Science, Mathematics, Engineering, and Technology Education*. Arlington, VA: National Science Foundation.
- Stewart, D.W. & Kamins, M.A. (1993). *Secondary Research: Information Sources and Methods* (2<sup>nd</sup> ed.). Newbury Park, CA: Sage Publications.
- Stewart, D.W. & Shamdasani, P.N. (1990). *Focus Groups: Theory and Practice*. Newbury Park, CA: Sage Publications.
- Strunk, W., Jr., & White, E. B. (1979). *The Elements of Style* (3<sup>rd</sup> ed.). New York, NY: Macmillan Publishing Co., Inc.
- Sylvia, R.D., Meier, K.J., & Gunn, E.M. (1985). *Program Planning and Evaluation for the Public Manager*. Monterey, CA: Brooks/Cole.
- U.S. Department of Defense. (March, 1996). *Key Criteria for Performance Measurement: An Action Plan for U.S. Department of Defense*. <http://www.Magi.com/~godbout/Kbase/metra5.htm>. Last accessed December 2000.
- U.S. General Accounting Office (July, 1995). *Financial Reporting: Checklist for Reports Prepared Under the CFO Act*. Publication No. GAO/AIMD-12.19.5B. Washington, DC.

- U.S. General Accounting Office (November, 2000). *Information Resources on Program Evaluation*. Center for Evaluation Methods and Issues. Washington, DC.
- U.S. General Accounting Office (1998). *Performance Measurement and Evaluation: Definitions and Relationships*. Publication No. GAO/GGD-98-26. Washington, DC.
- U.S. General Accounting Office. (1998). *Program Evaluation: Agencies Challenged by New Demand for Information on Program Results*. Publication No. GAO/GGD-98-53. Washington, DC.
- U.S. General Accounting Office. (1993). *Public Health Service: Evaluation Set-Aside Has Not Realized Its Potential to Inform the Congress..* Publication No. GAO/PEMD-99-13. Washington, DC.
- Veney, J.E., & Kaluzny, A.D. (1998). *Evaluation and Decision Making for Health Services* (3<sup>rd</sup> ed.). Chicago, IL: Health Administration Press.
- W.K. Kellogg Foundation. (1998). *Evaluation Handbook*. Battle Creek, MI.  
<http://www.wkkf.org/Publications/evalhdbk/>. Last accessed December 2000.
- Wholey, J.S. (1983). *Evaluation and Effective Public Management*. Boston, MA: Little, Brown and Company.
- Yin, R. K. (1989). *Case Study Research: Design and Methods*. Newbury Park, CA: Sage Publications.



## GLOSSARY

---

### *Archival Data*

Archival data (also called secondary data) is information collected prior to the evaluation and for another purpose. An extensive amount of archival data is available from NIH and/or other organizations, some of which may be relevant to the proposed evaluation. Typical archival data sources include program documents, computerized information systems, web sites, reports and other publications, and CD-ROMs containing census data and other national survey statistics. Commonly used strategies for collecting archival data include: document reviews, database extractions, web site reviews, and literature reviews. Using archival data generally requires much less time and expense than collecting new data.

### *Codebook*

A codebook documents how raw data will be synthesized, categorized, and transformed (usually to numeric values) so that the information gathered can be tabulated and analyzed using statistical tests and/or other standardized procedures.

### *Comparison Measures*

Comparison measures are measurements against which the performance of a program will be compared. There are three types of comparison measures: (1) measures of the program's prior performance, (2) measures of a comparable program's performance or a control group's performance, and (3) recognized standards of performance. See Appendix B for more information on comparison measures, including specific examples.

### *Conceptual Framework*

A conceptual framework describes, usually in the form of a diagram, how a particular program is intended to work. Specifically, it demonstrates how program resources, population characteristics, program activities, and external factors (if any) are expected to influence the achievement of a program's specific process, intermediate, and/or long-term goals. A conceptual framework (also called a logic model) may be simple or elaborate, and it can be developed using either common sense or a specific theory as a foundation. See Appendix D for examples of different types of conceptual frameworks.

### ***Descriptive Statistics***

Descriptive statistics are used to tabulate, depict, and describe collections of data. The data may be either quantitative (e.g., number of scientific papers published) or categorical (e.g., gender, geographic region) in nature. Examples of descriptive statistics include frequency distributions, contingency or cross-tabulation tables, measures of central tendency (e.g., mean, mode, median), and measures of variability (e.g., standard deviation, margin of error). These varied descriptive statistical techniques may be used to reduce a large amount of information into a more manageable (i.e., summarized) form.

### ***Direct Labor Costs***

Direct labor costs are expenses directly attributable to the workforce conducting the evaluation, based on the salary levels of the individuals performing the work. The mixture of skills, training, and experience needed by the individuals conducting the evaluation is often called the “labor mix.” Total direct labor costs depend on the labor mix as well as the hourly rate and total hours of effort for each labor category. Typically, the first step in constructing an evaluation budget estimate is to predict the mixture of skills needed to perform the project. See Appendix E for more information, including a sample evaluation budget.

### ***Efficiency Measures***

Efficiency measures are measurements of program performance that focus on the cost (in terms of dollars, FTEs, employee-hours, facilities, or other resources) per unit of output or outcome. Efficiency measures are sometimes called cost-effectiveness measures. See Appendix B for more information, including examples of efficiency measures (examples 6 and 7).

### ***Evaluation***

An evaluation is a formal appraisal of an object, which involves a systematic investigation of its worth and/or performance. In the case of a program evaluation, the object being investigated is a program (i.e., a set of activities designed to achieve one or more predefined goals). See Program and Program Evaluation.

### ***Expert Panels***

Expert panels are groups of individuals with expertise in specific areas. Meetings of expert panels usually involve a trained facilitator to stimulate discussion and help the group reach consensus, if possible, on a few major issues. For program evaluations, the issues to be considered by the panel are usually programmatic in nature and may involve making recommendations to program administrators. Expert panel meetings are sometimes audiotaped (with the approval of the panel members).

***External Factors***

External factors (sometimes called confounding variables) are conditions or circumstances beyond the control of the program that may influence program success. Examples include other programs with similar goals, unexpected positive events (such as a state budget surplus), and unexpected negative events (such as a natural disaster).

***Feasibility Study***

A feasibility study is a systematic assessment of the optimal approach for evaluating a program, including which evaluation designs and data collection strategies can and should be used. It usually includes determining whether conducting an evaluation is appropriate, designing a process evaluation or outcome evaluation for a proposed or existing program, and/or determining whether the evaluation can be conducted at a reasonable cost. A feasibility study may serve as a preliminary evaluation aimed at determining the optimal approach for a full-scale outcome evaluation. This type of study is sometimes called an evaluability assessment. See Appendix A for more information, including typical evaluation strategies.

***Fee***

The fee (or profit) in an evaluation budget estimate is the dollar amount over and above allowable costs that is to be paid to the organization responsible for conducting the evaluation. The complexity of the task, the level of risk to the organization performing the work, and other factors determine the fee, which is usually presented as a percentage of the total estimated costs. See Appendix E for more information, including a sample evaluation budget.

***Focus groups***

Focus groups are group interviews in which a trained facilitator asks general questions about one or more topics and encourages the participants to interact and consider each other's comments. The combined effort of the group may produce a wide range of information, insight, and ideas. Focus group sessions are often audiotaped, videotaped, and/or observed by others (with the approval of the focus group participants). See Paperwork Reduction Act Requirements.

***Hypothesis***

A hypothesis is an assumption about the relationship between two or more measurable variables. Inferential statistics may be used to test whether a specific hypothesis is true, assuming a predefined probability (significance level) that the hypothesized event could occur by chance. Study questions are frequently addressed by testing one or more hypotheses.

***Indirect Costs***

Indirect costs are expenses that are difficult to assign to specific project functions. They typically include: (1) fringe benefits for the individuals performing direct labor, (2) overhead costs, and (3) general and administrative (G&A) expenses. *Fringe benefits* may include paid holidays, vacations, sick leave, retirement benefits, and social security tax funded by the employer. *Overhead costs* generally include infrastructure expenses such as building rent and maintenance, utilities, and depreciation of equipment. *G&A expenses* often include the costs of personnel who are indirectly involved with project work (e.g., senior executives, human resource and accounting personnel) and other indirect costs (e.g., advertising, marketing, taxes). See Appendix E for more information, including a sample evaluation budget.

***Inferential Statistics***

Inferential statistics are used to make inferences (i.e., to draw conclusions or to generalize) about the properties of a population by examining a sample of the population. Inferential statistics are essential when assessing the relationships among program variables and testing hypotheses associated with particular study questions. Examples of inferential statistical methods include analysis of variance, regression analysis, correlation analysis, discriminant analysis, and analyses using chi-square tests, t-tests, and other parametric or nonparametric statistical tests.

***Intermediate Goals***

Intermediate goals describe specific outcomes the program should achieve in the near term. Examples of intermediate goals include increased publications in peer-reviewed journals, more individuals obtaining doctoral degrees in health-related sciences, development of an instrument for use in research or medicine that meets certain standards, and achievement of a specified level of satisfaction reported by scientists using the program. Intermediate goals are used primarily in outcome evaluations. See Appendix B for additional information.

***Inter-Rater Reliability*** – See Reliability.

***Institutional Review Board (IRB) Approval***

Studies that involve a potential risk to the rights and welfare of human subjects may require prior approval by an Institutional Review Board (IRB) of the study design, including the method for obtaining informed consent. Unlike clinical research studies, IRB approval is usually not required for program evaluations if potential respondents are clearly informed that they may choose not to participate.

***Logic Model*** – See Conceptual Framework. Also see Appendix D for specific examples.

### ***Long-Term Goals***

Long-term goals describe the ultimate outcomes the program is designed to achieve. Examples of long-term goals include discovery of a new treatment for a specific disease, more NIH-sponsored trainees/fellows pursuing biomedical research careers, and development of an improved approach for preventing disease or disability. Long-term goals are used primarily in outcome evaluations. See Appendix B for additional information.

### ***Needs Assessment***

A needs assessment is a type of program evaluation aimed at systematically determining the nature and extent of the problems that a proposed or existing program should address. It usually includes assessing the needs of stakeholders, developing appropriate program goals, and determining how a program should be designed or modified to achieve those goals. A needs assessment is often used as a tool for strategic planning and priority setting. See Appendix A for more information, including typical evaluation strategies.

### ***New Data***

New data (also called primary data) is information collected specifically for the evaluation. Commonly used strategies for collecting new data include personal interviews, focus groups, expert panels, questionnaires or other data collection instruments (i.e., forms) to be completed, adding evaluation questions to broader surveys (sometimes called omnibus surveys), and structured observations of program processes. Collecting new data generally requires more time and expense than using archival data. See Paperwork Reduction Act Requirements.

***OMB Clearance*** – See Paperwork Reduction Act Requirements.

### ***Other Direct Costs***

Other direct costs are expenses directly attributable to the proposed evaluation, excluding direct labor costs. Examples include costs related to consultants, subcontracts, meetings/travel, and miscellaneous supplies and services. See Appendix E for more information, including a sample evaluation budget.

### ***Outcome Evaluation***

An outcome evaluation is a systematic assessment of program accomplishments and effects to determine the extent to which a program's intermediate and/or long-term goals have been achieved. It usually includes examining the relationship between program activities and their effects, both intended and unintended, to identify why some program variations or strategies worked better than others. See Appendix A for more information, including typical evaluation strategies.

***Outcome Measures***

Outcome measures are measurements of program performance that focus on the intermediate and/or long-term accomplishments and effects of the program. See Appendix B for more information, including examples of outcome measures (examples 3-5).

***Output Measures***

Output measures are measurements of program performance that focus on the number of program activities conducted or products produced. See Appendix B for more information, including examples of output measures (examples 1-2).

***Paperwork Reduction Act Requirements***

The Paperwork Reduction Act of 1985 administered by the U.S. Office of Management and Budget (OMB) requires all federal agencies to obtain OMB clearance prior to collecting the same information from ten or more nonfederal employees, a process that often requires 6 to 9 months. Additional information on OMB clearance may be obtained by contacting your P&E Officer and/or the DHHS website <http://www.hhs.gov/oirm/infocollect/>.

***Performance Measures***

Performance measures are measurements of program performance during a given time period. There are three types of performance measures: (1) output measures, (2) outcome measures, and (3) efficiency measures. See Appendix B for more information, including specific examples.

***Personal Interviews***

Personal interviews may be conducted via telephone or in-person, with one individual or a few individuals, for the purpose of collecting data needed to answer questions on one or more topics. An interview guide (or discussion guide) is usually used by the interviewer to ask specific questions, some of which may be followed by probes for additional information. The interviewer generally summarizes the answers of the respondent(s) either during the interview or immediately afterward. Personal interviews are sometimes audiotaped (with the approval of the respondents). See Paperwork Reduction Act Requirements.

### ***Pilot Tests***

Pilot tests (sometimes called pretests) are trial runs designed to improve data collection instruments and procedures before the data collection effort is begun. They are usually conducted as part of a feasibility study. Pilot tests typically include (1) using the data collection instruments to examine a small number of cases (e.g., asking a few individuals to fill out questionnaires, conducting interviews with a few people, completing a few observations, examining a small set of records); (2) reviewing the completed forms for problem areas (e.g., blank responses, misinterpretations); (3) conducting personal or group interviews with the data collectors and/or respondents to discuss their general impressions of the questionnaire and to identify any items that were difficult to understand or problematic, (4) analyzing the pilot data collected to determine the effectiveness of the instruments in gathering the desired information, (5) using the analyses and comments to revise the data collection instruments and procedures, and (6) conducting field tests of the data collection instruments and procedures to find out how they work under realistic conditions. Field tests are particularly useful for determining the overall feasibility of the proposed data collection and analysis strategies, making final revisions, and estimating the total costs of the study.

### ***Population Characteristics***

Population characteristics are variables that describe differences among the members of the target population, particularly characteristics that may be related to program success. Examples include demographic characteristics (e.g., age, gender, socioeconomic status), measures of health status, and characteristics of grant applications received by an IC during a given period.

### ***Privacy Act Requirements***

The Privacy Act of 1974 restricts the use and disclosure of personally identifiable information maintained by NIH and other federal agencies in organized “systems of records.” The Privacy Act also specifies that information collected for one purpose may not be used for another purpose without notifying or obtaining the consent of the subject of the record. In program evaluations, the Privacy Act generally applies when data to be collected and maintained can be linked with a personal identifier (e.g., name, Social Security Number, date of birth, patient identifier, or a randomly assigned computer number that is linked to a master index of individual identifiers). In cases where the Privacy Act applies, a Privacy Act Notification Statement is required so that potential study participants know: (1) the statutory authority for the data collection; (2) whether or not their response is voluntary; (3) the consequence, if any, of not providing the information; and (4) the extent to which confidentiality of the information is protected. Additional information on the Privacy Act may be obtained by contacting the NIH Privacy Act Officer and/or the NIH website <http://oma.od.nih.gov/ms/privacy/>.

### ***Process Evaluation***

A process evaluation is a systematic assessment of program operations to determine whether a program is being conducted as planned, whether expected output is being produced, and/or how program-critical processes can be improved. It usually includes assessing the extent to which process goals have been achieved. See Appendix A for more information, including typical evaluation strategies.

### ***Process Goals***

Process goals describe how the program should operate and what levels of output should be expected. Process goals are often expressed in terms of the number of activities to be conducted, services to be provided, products to be produced, or efficiency of program operations to be achieved during a given time period. Examples of process goals include adherence to a pre-established timeline and budget, an increased level of program activities, and a reduction in unit costs. Process goals are used primarily in process evaluations, although they may also be used in outcome evaluations. See Appendix B for additional information.

### ***Program***

For the purposes of NIH One Percent Evaluation Set-Aside Funds, the term “program” is broadly defined as a set of activities funded by NIH to achieve one or more predefined goals (referred to as “program goals”). Examples of programs include national health awareness campaigns, initiatives to enhance the research capacity of academic institutions, grants management programs, training programs for intramural researchers, and activities to improve the efficiency and/or effectiveness of NIH operations (e.g., computerized information systems, web sites).

### ***Program Activities***

Program activities are the specific actions, operations, processes, or other functions that are essential to the conduct of the program. Examples include initiating pilot projects, holding workshops, reviewing grants, holding media events, and providing new incentives to encourage research.

### ***Program Evaluation***

Program evaluations are systematic investigations or studies that involve assessing the worth and/or performance of particular programs. In most cases, the underlying purpose of a program evaluation is to help program administrators improve a program or make other programmatic decisions. There are four types of program evaluations: (1) needs assessments, (2) feasibility studies, (3) process evaluations, and (4) outcome evaluations. Needs assessments and feasibility studies are usually conducted as preliminary studies to improve the design of a more complex process evaluation or outcome evaluation. Experts external to the program often conduct program evaluations, but program managers may also conduct them. Additional information is presented in Appendix A.



### ***Program Goals***

Program goals are the intended effects of a program, as noted in authorizing legislation or other documents written when the program was established. For a program that is not yet established, the proposed program goals should summarize the intended effects of the new program. There are three types of program goals: (1) process goals, (2) intermediate goals, and (3) long-term goals. The extent to which a particular program goal has been achieved should be assessed using one or more performance measures and comparison measures (as described in Appendix B).

### ***Program Resources***

Program resources are the funding, human capital (e.g., FTEs), infrastructure, and/or other assets allocated to the program or specific program components during a given time period. Examples include employee-hours, total square feet of laboratory space, and the average amount of dollars budgeted per year for program salaries and wages, consultant services, equipment, supplies, travel costs, and other direct costs.

### ***Qualitative Analyses***

Qualitative analyses are used to describe and/or interpret data presented in the form of words rather than numbers. In program evaluations, qualitative analyses are typically used when data are collected from document reviews, expert panels, focus groups, personal interviews, and structured observations. Examples include pattern (or thematic) coding, content analysis, triangulation, within-case and cross-case analyses, and the use of matrices, chronological models, and other displays to explain qualitative findings.

### ***Quality Control Procedures***

Quality control procedures refer to the steps taken to improve the reliability and validity of the data collected and analyzed in a program evaluation. The most common quality control procedure is checking the data for inconsistent, unlikely, or otherwise erroneous responses. Other commonly used quality control procedures include training and monitoring of individuals handling data (e.g., data collectors, coders, data entry personnel, and data analysts), developing written instructions and codebooks, and conducting pilot tests of instruments and procedures, inter-rater reliability checks, and double data entry.

### ***Quantitative Analyses***

Quantitative analyses are used to describe and/or interpret data presented in the form of numbers rather than words. This type of data can be measured along a continuum and is characterized by having additive properties, equal intervals, and usually a zero point. Both descriptive and inferential statistical techniques may be used with quantitative data.

## ***Questionnaires***

Questionnaires are written data collection instruments (i.e., forms) that include instructions and a set of questions about one or more topics. They may be administered in person, by mail, or electronically (e.g., via e-mail or web sites). Newly developed questionnaires should be pilot tested for effectiveness. See Paperwork Reduction Act Requirements.

## ***Reliability***

Reliability is the extent to which a data collection instrument or effort yields consistent and stable results over repeated measurements conducted under similar conditions. For example, a bathroom scale is unreliable if it produces three different weights in three consecutive weighings of the same person. Reliability may be assessed in several ways, including (1) *inter-rater reliability*, which measures the similarity of scores assigned by different raters (e.g., interviewers, observers) to the same phenomenon; (2) *test-retest reliability*, which measures the similarity of scores (or responses to a particular set of questions) obtained at different times from the same individuals; and (3) *internal-consistency reliability* (sometimes called split-halves reliability), which measures the similarity of scores obtained from the same individuals responding to two sets of questions designed to measure the same concept, construct, or trait. A data collection instrument or effort must be reliable to be considered valid.

## ***Response Rate***

The response rate for a data collection effort is the number of actual respondents divided by the number of potential respondents. The denominator consists of all of the individuals in the target population who were sampled, including those who did not respond for a particular reason (e.g., refusal, language problems, inability to contact).

## ***Sample***

A sample is a subset of individuals or objects selected (or drawn) from the target population by means of a sampling strategy.

## ***Sample Size***

Sample size is the number of individuals or objects selected from the target population for data collection purposes. The size of a sample is important because it must be large enough to make valid inferences about the population from which the sample was drawn. Many factors should be considered when determining the sample size, including the planned sampling strategy, the number of subgroups within the target population for which separate estimates are required, and estimates of the proportion of the population that fall into those subgroups.

### ***Sampling Strategy***

A sampling strategy is the approach used to select a sample. Sampling strategies are used to increase the likelihood that the inferences made about the target population are valid. Examples include one-stage sampling techniques (e.g., simple random sampling, stratified sampling) or multi-stage sampling techniques (e.g., random digit dialing, area probability sampling).

### ***Stakeholders***

Stakeholders are individuals or groups who are likely to be interested in, to be impacted by, or to use the findings of the evaluation. Stakeholders typically include those involved in program operations (e.g., NIH researchers, staff, administrators), those served by the program, and others who have an investment or interest in the program.

### ***Standards of Performance***

Standards of performance are levels of program processes, outputs, and/or outcomes established by authority or general agreement as being acceptable. Examples include a defined timeline or budget, a certain level of work output or product/service quality, and a specific outcome.

### ***State-of-the-Science Assessment***

A state-of-the-science assessment is a systematic review of existing research and recent advances in a specific area of biomedical research for the purpose of identifying scientific achievements, gaps, and opportunities. It is usually conducted via a conference, workshop, or expert panel meeting. State-of-the-science assessments are designed to help NIH program administrators and researchers identify research priorities, and develop or modify program goals.

### ***Strategic Planning***

Strategic planning is a process that involves setting goals for a program or organization, developing strategies for achieving those goals, and determining how success will be measured and evaluated.

### ***Structured Observation***

A structured observation is a type of data collection in which the situation of interest is watched by one or more observers trained to record relevant facts, actions, and behaviors in a standardized way. Structured observations are usually recorded on data collection forms and may include the use of audiotape or videotape (with the approval of the individuals being observed). See Paperwork Reduction Act Requirements.

### ***Study Questions***

Study questions are the key questions that the evaluation is designed to answer. For process evaluations and outcome evaluations, the study questions usually address the extent to which specific program goals have been achieved. Study questions are often answered by testing specific hypotheses. Examples of study questions are presented in Appendix C.

### ***Target Population***

The target population is the primary group about which information is needed to answer the study questions. It is frequently a group of individuals having certain characteristics, such as the participants in a specific NIH training program, the members of an IC's scientific review groups (study sections), the individuals who called an NIH health hotline during a given time period, or the NIH administrators who implemented a new program. The target population may also consist of a group of objects having certain characteristics, such as the academic institutions funded or the R01 grants awarded by an IC during a given period.

### ***Unit of Analysis***

The unit of analysis is the individual item within the target population for which data will be collected and analyzed to answer the study questions. The unit of analysis, for example, may be defined as a program participant, a member of the general public who accessed an NIH service, an academic department, or an individual grant award. In some cases, more than one unit of analysis may be included in the evaluation design.

### ***Validity***

Validity is the extent to which a data collection instrument or effort accurately measures what it is supposed to measure. Validity may be assessed in several different ways, for example: (1) the *face validity* of a questionnaire or other data collection instrument is assessed using human judgment, frequently the judgment of a group of experts in the field, to determine whether the instrument appears to measure what it claims to measure; (2) *construct validity* is assessed by determining the extent to which the underlying construct, concept, or theory accounts for respondents' scores; (3) *concurrent validity* is assessed by comparing the similarity of respondents' scores to other criteria that are assumed to measure the same construct; and (4) *predictive validity* is assessed by comparing respondents' scores to future measures of performance. It is generally agreed that there is no simple, uniform, wholly objective procedure for determining the validity of a data collection instrument or effort. For a data collection instrument or effort to be considered valid, it must also be reliable.

***Variable***

A variable is a factor, construct, or characteristic of a person, object, or program that can take on different values. In a program evaluation, the key variables are those for which data will be gathered to answer one or more study questions. Measures of a program's performance (e.g., output measures, outcome measures, or efficiency measures) are sometimes called *dependent variables*, while factors that may be predictive of a program's performance (e.g., program resources, population characteristics, or program activities) are sometimes called *independent variables*.